

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-039347

(43)Date of publication of application : 12.02.1999

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-213993

(71)Applicant : NEC CORP

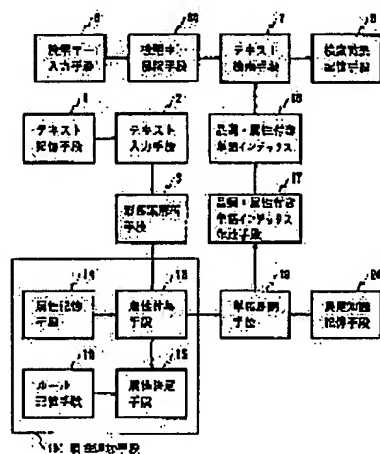
(22)Date of filing : 24.07.1997

(72)Inventor : TAKEMOTO YOSHIMI

(54) TEXT RETRIEVAL SYSTEM, INDEX GENERATION DEVICE, TEXT RETRIEVAL DEVICE AND RECORDING MEDIUM WHICH COMPUTER CAN READ**(57)Abstract:**

PROBLEM TO BE SOLVED: To suppress retrieval excess and to realize retrieval by means of the synonym of different marking.

SOLUTION: At the time of generating a word index 18 from a text, not only information of the text containing the index of a word and the text containing the word but also information of a part of speech (proper noun on 'USA', for example) obtained by the analysis of morpheme, and the attribute of the word added by an attribute addition means 100 (country name on 'USA', for example) are registered. A polygamous word ('America' on 'USA', for example) is developed and registered based on the part of speech and the attribute of the word. When a retrieval key ('America', for example) is inputted at the time of retrieval, a text retrieval means 7 retrieves the word index 18 and information on the text containing not only 'America' but also 'USA' being the synonym is obtained. When the retrieval key with the part of speech and the attribute as retrieval conditions is inputted, the text retrieval means 7 retrieves information of the text containing the word satisfying information of the word and the part of speech or the attribute from the word index 18.



BEST AVAILABLE COPY

LEGAL STATUS

[Date of request for examination]

24.07.1997

[Date of sending the examiner's decision of rejection]

06.06.2000

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-39347

(43) 公開日 平成11年(1999) 2月12日

(51) Int.Cl.⁶

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/40

15/401

3 7 0 A

3 1 0 A

審査請求 有 請求項の数10 F D (全 15 頁)

(21) 出願番号 特願平9-213993

(22) 出願日 平成9年(1997) 7月24日

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 竹元 義美

東京都港区芝五丁目7番1号 日本電気株式会社内

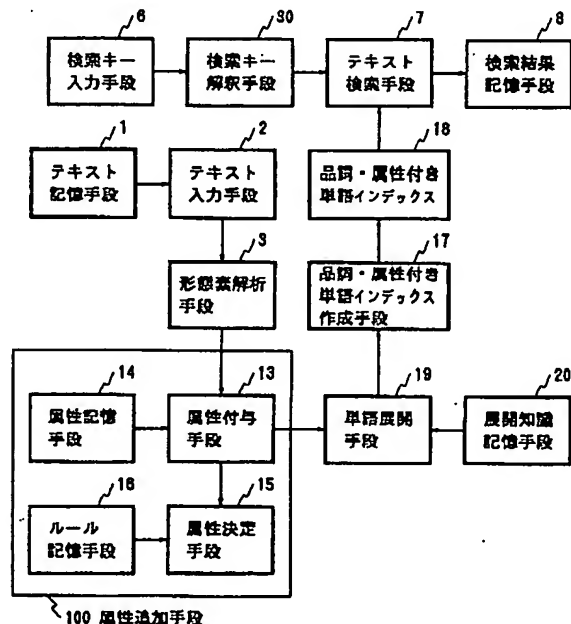
(74) 代理人 弁理士 境 廣巳

(54) 【発明の名称】 テキスト検索システム、インデックス作成装置、テキスト検索装置及びコンピュータ読み取り可能な記録媒体

(57) 【要約】

【課題】 検索過剰を抑制しつつ、異表記の同義語による検索も可能にする。

【解決手段】 テキストから単語インデックス18を作成する際、単語の見出しとその単語を含むテキストの情報だけでなく、形態素解析で得られたその単語の品詞（例えば「米」について固有名詞）、属性追加手段100で追加されたその単語の属性（例えば「米」について国名）の情報を登録する。また、単語の品詞や属性に基づき多義性のある語（例えば「米」について「アメリカ」）を展開して登録する。検索時、検索キー（例えば「アメリカ」）が入力されると、テキスト検索手段7はその単語で単語インデックス18を検索し、「アメリカ」のみならずその同義語である「米」を含むテキストの情報も得る。また品詞や属性を検索条件とした検索キーが入力されると、テキスト検索手段7は、単語と品詞または属性の情報との双方を満たす単語を含むテキストの情報を単語インデックス18から検索する。



【特許請求の範囲】

【請求項 1】 検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段と、検出された単語及び品詞とその単語が出現するテキストの情報とを、単語インデックスに登録する単語インデックス作成手段とを備えるインデックス作成装置。

【請求項 2】 検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段と、検出された単語に対し品詞以外の属性を決定して付与する属性追加手段と、検出された単語、品詞、属性およびその単語が出現するテキストの情報を、単語インデックスに登録する単語インデックス作成手段とを備えるインデックス作成装置。

【請求項 3】 検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段と、検出された単語に対し品詞以外の属性を決定して付与する属性追加手段と、検出された単語の品詞及び属性を考慮して、検出された単語を同じ意味の異なる表記の単語に展開する単語展開手段と、検出された単語、品詞、属性およびその単語が出現するテキストの情報と、展開して得られた単語、展開元の単語の品詞、属性およびその単語が出現するテキストの情報とを、単語インデックスに登録する単語インデックス作成手段とを備えるインデックス作成装置。

【請求項 4】 前記属性追加手段は、単語の品詞情報とその属性情報とを記憶する属性記憶手段と、形態素解析で検出された単語に対し、検出された品詞の情報と前記属性記憶手段の記憶内容とに基づき属性を付与する属性付与手段と、単語列に現れる品詞または属性の並びのパターン等に基づき単語の属性を決定するルールを記憶するルール記憶手段と、該ルール記憶手段に記憶されたルールを適用して前記属性付与手段で付与された属性を検査し必要に応じて修正する属性決定手段とを含むことを特徴とする請求項 2 または 3 記載のインデックス作成装置。

【請求項 5】 指定された単語を含むテキストを検索するテキスト検索システムにおいて、請求項 1、2、3 または 4 の何れか 1 項に記載されたインデックス作成装置と、品詞または属性付きの検索キーを、単語とその品詞または属性の情報とに解釈する検索キー解釈手段と、解釈された単語と品詞または属性の情報との双方を満たす単語を含むテキストの情報を前記インデックス作成装置で作成された単語インデックスから検索するテキスト検索手段とを含むテキスト検索装置とを有することを特徴とするテキスト検索システム。

【請求項 6】 単語とその品詞または属性の候補とを記憶する候補記憶手段と、ユーザから入力された検索キーにかかる単語の属性候補を前記候補記憶手段から検索し

て表示し、その内の一つをユーザに選択させる選択手段と、ユーザから入力された検索キーにかかる単語と前記選択された属性の情報との双方を満たす単語を含むテキストの情報を単語インデックスから検索するテキスト検索手段とを含むテキスト検索装置。

【請求項 7】 指定された単語を含むテキストを検索するテキスト検索システムにおいて、請求項 1、2、3 または 4 の何れか 1 項に記載されたインデックス作成装置と、

10 請求項 6 に記載されたテキスト検索装置とを有するテキスト検索システム。

【請求項 8】 コンピュータを、検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段、検出された単語及び品詞とその単語が出現するテキストの情報とを、単語インデックスに登録する単語インデックス作成手段、として機能させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

20 【請求項 9】 コンピュータを、検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段、検出された単語に対し品詞以外の属性を決定して付与する属性追加手段、検出された単語、品詞、属性およびその単語が出現するテキストの情報を、単語インデックスに登録する単語インデックス作成手段、として機能させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

30 【請求項 10】 コンピュータを、検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段、検出された単語に対し品詞以外の属性を決定して付与する属性追加手段、検出された単語の品詞及び属性を考慮して、検出された単語を同じ意味の異なる表記の単語に展開する単語展開手段、検出された単語、品詞、属性およびその単語が出現するテキストの情報と、展開して得られた単語、展開元の単語の品詞、属性およびその単語が出現するテキストの情報とを、単語インデックスに登録する単語インデックス作成手段、として機能させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、指定された単語を含むテキストを検索するテキスト検索システムに関する。

50 【0002】

【従来の技術】大量のテキストを対象に、指定された単語を含むテキストを検索する手法の一つとして広く用いられているものに、単語単位の見出しにその単語を含むテキストの文書番号情報を持つインデックスを利用する技術（例えば特開平7-182370号公報）がある。

【0003】図18は従来のテキスト検索システムの構成を示すブロック図である。同図に示すように、従来のテキスト検索システムは、テキスト記憶手段1、テキスト入力手段2、形態素解析手段3、単語インデックス作成手段4、単語インデックス5、検索キー入力手段6、
10 テキスト検索手段7および検索結果記憶手段8で構成されている。このうち、テキスト記憶手段1、テキスト入力手段2、形態素解析手段3および単語インデックス作成手段4が、インデックス作成装置を構成し、テキスト記憶手段1に記憶されたテキストを処理して単語インデックス5を作成する。検索キー入力手段6、テキスト検索手段7および検索結果記憶手段8で構成されるテキスト検索装置は、上記作成された単語インデックス5を利用してテキスト検索を行う。

【0004】形態素解析手段3は、テキスト記憶手段1
20 に格納されたテキストをテキスト入力手段2から入力し、形態素解析を実行する。ここでの形態素解析は、入力されたテキストを単語辞書を用いて単語単位に認定する解析処理であり、テキスト処理によく用いられている。なお、形態素解析技術の詳細については、「国語辞書の記憶と日本語文の自動分割」（長尾他、情報処理Vol. 19, No. 6, 1978年）などに記載されている。

【0005】単語インデックス作成手段4は、形態素解析手段3で認定された単語を見出しとして、その単語が
30 どのテキストに含まれるかの情報（例えばテキスト識別子やそれに加えてテキスト中での位置情報）を付与した単語インデックス5を作成する。

【0006】図19に単語インデックスの内容例を示す。同図に示すように、従来の単語インデックスは、テキスト中に出現した単語を見出しとし、その単語の出現したテキストの情報（図の例ではテキスト識別子）が検索できるようにしている。

【0007】検索キー入力手段6はユーザが検索キーを入力するための手段、テキスト検索手段7は検索キー入力手段6により入力された検索キーを含むテキストを単語インデックス5から検索する手段、検索結果記憶手段8は検索結果を格納する手段である。

【0008】

【発明が解決しようとする課題】上述したように、従来の単語インデックスは、テキスト中に出現した単語そのものとテキスト情報との関係を保持しているに過ぎないため、検索キーと同じ表記であれば意味が異なる単語でもヒットしてしまい、検索過剰を生じる問題がある。例えば、「アメリカ」という意味で「米」を検索した場合、

「コメ」という意味の「米」もヒットしてしまい、検索過剰となる。

【0009】この種のテキスト検索システムについては従来より各種の改良技術が提案されているが、その多くは検索漏れの解消にあり、検索過剰は余り問題としていない。例えば特開平7-152778号公報では、インデックス中のキーワードとは別にその上位概念見出しを設け、ユーザの入力したキーワードが上位概念見出しと一致するか否かを調べている。また、特開平9-44521号公報では、複数の異なる表記が用いられる語については、それら異なる表記の語に共通な識別子に対応させてインデックスに登録しておき、検索時にユーザの入力した語をその語に対応する識別子に変換し、識別子でインデックスを検索している。何れも、検索漏れは少なくなるが、検索過剰は却って目立つようになる。

【0010】本発明の目的は、検索キーと同じ表記でも意味の異なる単語は検索されないようにして、検索過剰をできるだけ抑制することにある。

【0011】

【課題を解決するための手段】本発明は上記の目的を達成するために、以下のインデックス作成装置を提供する。

【0012】第1のインデックス作成装置は、検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段と、検出された単語及び品詞とその単語が出現するテキストの情報とを、単語インデックスに登録する単語インデックス作成手段とを備えている。

【0013】第2のインデックス作成装置は、検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段と、検出された単語に対し品詞以外の属性を決定して付与する属性追加手段と、検出された単語、品詞、属性およびその単語が出現するテキストの情報を、単語インデックスに登録する単語インデックス作成手段とを備えている。

【0014】第3のインデックス作成装置においては、検索対象となるテキストを形態素解析し、テキスト中に含まれる単語とその品詞とを検出する形態素解析手段と、検出された単語に対し品詞以外の属性を決定して付与する属性追加手段と、検出された単語の品詞及び属性を考慮して、検出された単語を同じ意味の異なる表記の単語に展開する単語展開手段と、検出された単語、品詞、属性およびその単語が出現するテキストの情報とを、単語インデックスに登録する単語インデックス作成手段とを備えている。

【0015】なお、第2および第3のインデックス作成装置における属性追加手段は、たとえば、単語の品詞情報とその属性情報とを記憶する属性記憶手段と、形態素

解析で検出された単語に対し、検出された品詞の情報と前記属性記憶手段の記憶内容とに基づき属性を付与する属性付与手段と、単語列に現れる品詞または属性の並びのパターン等に基づき単語の属性を決定するルールを記憶するルール記憶手段と、該ルール記憶手段に記憶されたルールを適用して前記属性付与手段で付与された属性を検査し必要に応じて修正する属性決定手段とで構成される。

【0016】また、上記の各インデックス作成装置と組み合わせられるテキスト検索装置として、本発明は以下のようなテキスト検索装置を提供する。

【0017】第1のテキスト検索装置は、品詞または属性付きの検索キーを、単語とその品詞または属性の情報とに解釈する検索キー解釈手段と、解釈された単語と品詞または属性の情報との双方を満たす単語を含むテキストの情報を単語インデックスから検索するテキスト検索手段とを備えている。

【0018】第2のテキスト検索装置は、単語とその品詞または属性の候補とを記憶する候補記憶手段と、ユーザから入力された検索キーにかかる単語の品詞または属性の候補を前記候補記憶手段から検索して表示し、その内の一つをユーザに選択させる選択手段と、ユーザから入力された検索キーにかかる単語と前記選択された品詞または属性の情報との双方を満たす単語を含むテキストの情報を単語インデックスから検索するテキスト検索手段とを備えている。

【0019】本発明のテキスト検索システムは、上記第1、第2、第3のインデックス作成装置の何れか一つと、上記第1、第2のテキスト検索装置の何れか一つとが組み合わされて構成される。

【0020】

【発明の実施の形態】次に本発明の実施の形態の例について図面を参照して詳細に説明する。

【0021】図1は本発明の第1の実施例のブロック図である。この例のテキスト検索システムは、テキスト記憶手段1、テキスト入力手段2、形態素解析手段3および品詞付き単語インデックス作成手段11から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞付き単語インデックス12、検索キー入力手段6、検索キー解釈手段30、テキスト検索手段7および検索結果記憶手段8から構成されるテキスト検索装置とを有している。

【0022】以下、インデックス作成装置とテキスト検索装置とに分けて説明する。

【0023】○インデックス作成装置

テキスト記憶手段1には、検索対象となる複数のテキストが記憶されている。テキスト入力手段2は、テキスト記憶手段1から一つずつテキストを読み込み、形態素解析手段3に入力する。形態素解析手段3は、入力されたテキストについて形態素解析を実行する。ここでの形態

素解析は、入力されたテキストを単語辞書を用いて単語単位に認定し、かつ各単語の品詞を認定する解析処理である。

【0024】品詞付き単語インデックス作成手段11は、形態素解析手段3の結果である単語と品詞の情報および今回処理されたテキストの識別子とを、品詞付き単語インデックス12に登録する。単語の品詞情報を追加した点が、図19で説明した従来の単語インデックス5との相違点である。

【0025】○テキスト検索装置

検索キー入力手段6は、ユーザが指定した検索キーを入力する。検索キーは、従来は単語のみで構成されていたが、本実施例では、単語に加えてその品詞を指定することができる。単語と品詞を指定した検索キーを品詞付き検索キーと呼ぶ。検索キー解釈手段30は、品詞付き検索キーが入力されると、それを解釈して、単語と品詞の情報とを認識し、テキスト検索手段7に伝達する。

【0026】テキスト検索手段7は、検索キー解釈手段30から渡された単語と品詞の情報との双方を満たす単語を、品詞付き単語インデックス12から検索し、それに対応するテキストの情報を検索結果記憶手段8に格納する。検索結果記憶手段8に検索結果を格納する代わりに、またそれに加えて、図示しない表示装置やプリンタに検索結果を出力しても良い。

【0027】品詞付き検索キーでなく、従来と同じく、単語のみからなる検索キーによる検索も可能である。この場合、テキスト検索手段7は品詞付き単語インデックス12の品詞の情報を無視し、従来と同様の検索を行う。

【0028】次に、具体例を挙げて本実施例の動作を説明する。

【0029】テキスト記憶手段1に記憶されているテキストが、図2に示すテキストA1、A2であるとする。このテキストA1、A2について、形態素解析手段3が形態素解析すると、例えば図3に示すような解析結果が得られる。図3では、テキストA1について、「大分」、「県」、「は」をそれぞれ単語と認定し、その品詞をそれぞれ固有名詞、固有名詞接辞、助詞としている。また、テキストA2について、「大分」、「昔」、「に」をそれぞれ単語と認定し、その品詞をそれぞれ副詞、名詞、助詞としている。この解析結果に基づき、品詞付き単語インデックス作成手段11は、図4に示すような情報を品詞付き単語インデックス12に登録する。

【0030】ユーザは、固有名詞（県名）である「大分」を含むテキストを検索したいとする。このときユーザは、単語「大分」と共にその品詞である固有名詞を指定した品詞付き検索キー「大分@固有名詞」を入力する。ここで、@は単語と品詞の情報を区切るマークである。検索キー解釈手段30は、@の前の部分を単語、@の後ろの部分を品詞の情報と解釈し、テキスト検索手

段7に伝達する。テキスト検索手段7は、単語「大分」をキーに図4に示す品詞付き単語インデックス12を検索すると、図2のテキストA1、A2が求まる。ここで、品詞を比較すると、テキストA1は合致するが、テキストA2の品詞は副詞で品詞付き検索キーで指定された固有名詞と異なるため、テキストA1のみを検索結果とする。

【0031】これに対し、図18で説明した従来のテキスト検索装置では、品詞による絞り込みができないため、テキストA1だけでなくテキストA2も検索結果に 10 含まれることになり、検索過剰となる。

【0032】図5は本発明の第2の実施例のブロック図である。この例のテキスト検索システムは、テキスト記憶手段1、テキスト入力手段2、形態素解析手段3、属性追加手段100および品詞・属性付き単語インデックス作成手段17から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス18、検索キー入力手段6、検索キー解釈手段30、テキスト検索手段7および検索結果記憶手段8から構成されるテキスト検索装置と 20 を有している。

【0033】本実施例は、各単語の品詞の情報に加え、品詞以外の属性を単語に追加することにより、より一層の絞り込みを可能としたものである。

【0034】以下、インデックス作成装置とテキスト検索装置とに分けて説明する。

【0035】○インデックス作成装置

テキスト記憶手段1には、検索対象となる複数のテキストが記憶されている。テキスト入力手段2は、テキスト記憶手段1から一つずつテキストを読み込み、形態素解析手段3に 30 入力する。形態素解析手段3は、入力されたテキストについて形態素解析を実行する。ここでの形態素解析は、入力されたテキストを単語辞書を用いて単語単位に認定し、かつ各単語の品詞を認定する解析処理である。

【0036】属性追加手段100は、形態素解析手段3で認定された単語に対し品詞以外の属性を決定して付与する。属性追加手段100は、属性付与手段13と属性記憶手段14と属性決定手段15とルール記憶手段16とから構成される。

【0037】属性記憶手段14は、所定の単語毎に、その単語の品詞と品詞以外のその単語の属性との関係を記憶している。例えば、「アメリカ」という単語について、品詞「固有名詞」と属性「国名」との対を記憶している。

【0038】属性付与手段13は、形態素解析手段3の解析結果の各単語に対して、若しその単語の属性が属性記憶手段14に記憶されていれば、その属性をその単語に付与する。このとき、解析結果の品詞と属性記憶手段14中の品詞とが完全に一致しない場合は、未決定の印 50

(例えば?)を付けておく。例えば、解析結果中の単語「アメリカ」が品詞「固有名詞」に認定されており、属性記憶手段14に「アメリカ」について品詞「固有名詞」と属性「国名」が記憶されていれば、単語「アメリカ」に属性「国名」を付与する。他方、解析結果中の単語「米」が品詞「名詞」に認定されており、属性記憶手段14に「米」について品詞「固有名詞」と属性「国名」が記憶されている場合、単語「米」に属性「国名?」を付与する。このように未決定の属性は、最終的に属性決定手段15でその可否が決定される。

【0039】ルール記憶手段16は、属性付与手段13によって付与された未決定の属性の妥当性を判断するためのルールを記憶する。属性決定手段15は、このルールを適用して、属性付与手段13によって付与された未決定の属性の妥当性を検査し、付与の可否を判断する。ルール記憶手段16に格納されるルールは、形態素解析手段3の結果得られる品詞や属性付与手段13によってその単語に付与された属性の並びのパターンなどに基づいて記述される。以下に、未決定の属性「国名?」に利用できるルールの例を示す。

【0040】①「人名」+「国名?」+「人名接辞」のパターンであれば、その「国名?」は「国名」である。

②「国名?」+「政治機関名」のパターンであれば、その「国名?」は「国名」である。

③「国名?」+「通貨単位」のパターンであれば、その「国名?」は「国名」である。

④「国名?」+「で」(助詞)+「は」(助詞)のパターンであれば、その「国名?」は「国名」である。

⑤「国名?」+「、」(読点)+「国名?」のパターンであれば、その「国名?」は「国名」である。

⑥「国名?」が連続(近接)するパターンであれば、その「国名?」は「国名」である。

【0041】以上のようなルールを利用すれば、属性決定手段15は、以下の未決定の属性「国名?」を「国名」に決定することができる。

(a)「クリントン(固有名詞:人名)/米(名詞:国名?)/大統領(名詞:人名接辞)」

(b)「米(名詞:国名?)/外務省(名詞:政治機関名)」

40 (c)「米(名詞:国名?)/ドル(接辞:通貨単位)」

(d)「米(名詞:国名?)/で(助詞)/は(助詞)/、(読点)」

(e)「米(名詞:国名?)/、(読点)/仏(名詞:国名?)/、(読点)/英(名詞:国名?)」

なお、括弧内で「:」より前の情報は形態素解析手段3の結果得られる品詞情報であり、「:」以下の情報が属性記憶手段14に記述されていて、属性付与手段13により付与された情報である。

【0042】品詞・属性付き単語インデックス作成手段 50

17は、形態素解析手段3の結果である単語と品詞の情報ならびに属性追加手段100で追加された属性と、今回処理されたテキストの識別子とを、品詞・属性付き単語インデックス18に登録する。品詞の情報に加え、更に品詞以外の属性を追加した点が、図1の実施例との相違点である。

【0043】○テキスト検索装置

検索キー入力手段6は、ユーザが指定した検索キーを入力する。検索キーは、従来は単語のみで構成されていたが、本実施例では、単語に加えてその品詞や属性を指定することができる。単語と品詞を指定した検索キーを品詞付き検索キーと呼ぶ。単語と属性を指定した検索キーを属性付き検索キーと呼ぶ。単語と品詞と属性を指定した検索キーを品詞・属性付き検索キーと呼ぶ。検索キー解釈手段30は、検索キーが入力されると、それを解釈して、単語と品詞と属性の情報とを認識し、テキスト検索手段7に伝達する。

【0044】テキスト検索手段7は、検索キー解釈手段30から単語と品詞の情報が渡されると、単語と品詞の情報の双方を満たす単語を、品詞・属性付き単語インデックス18から検索し、単語と属性の情報が渡されると、単語と属性の情報の双方を満たす単語を、品詞・属性付き単語インデックス18から検索し、単語と品詞と属性の情報が渡されると、単語と品詞と属性の情報の全てを満たす単語を、品詞・属性付き単語インデックス18から検索する。そして、それに対応するテキストの情報を検索結果記憶手段8に格納する。検索結果記憶手段8に検索結果を格納する代わりに、またそれに加えて、図示しない表示装置やプリンタに検索結果を出力するようにしても良い。

【0045】品詞や属性の付いた検索キーでなく、従来と同じく、単語のみからなる検索キーによる検索も可能である。この場合、テキスト検索手段7は品詞・属性付き単語インデックス18の品詞、属性の情報を無視し、従来と同様の検索を行う。

【0046】次に、具体例を挙げて本実施例の動作を説明する。

【0047】テキスト記憶手段1に記憶されているテキストが、図6に示すテキストB1、B2、B3であるとする。このテキストB1、B2、B3について、形態素解析手段3が形態素解析すると、例えば図7に示すような解析結果が得られる。図7では、テキストB1について、「アメリカ」、「は」をそれぞれ単語と認定し、その品詞をそれぞれ固有名詞、助詞としている。また、テキストB2について、「日」、「米」、「間」、「の」、「貿易」をそれぞれ単語と認定し、その品詞をそれぞれ名詞、名詞、名詞、助詞、サ変名詞としている。更に、テキストB3について、「新潟」、「産」、「の」、「米」、「を」それぞれ単語と認定し、その品詞をそれぞれ固有名詞、名詞、助詞、名詞、助詞として

いる。

【0048】属性記憶手段14に、図8に示すような属性情報が格納されているとすると、属性付与手段13は、図7の形態素解析結果中の単語「アメリカ」、「日」、「米」、「新潟」に対して図9に示すような属性情報を付与する。ここで、テキストB2中の「日」、「米」と、テキストB3中の「米」とには、解析された品詞が名詞、属性記憶手段14中の品詞が固有名詞であるため、未決定の属性「国名？」が付与される。

10 【0049】ルール記憶手段16に、前述した①～⑥のルールが格納されているとすると、属性決定手段15は、テキストB2中の「日」、「米」の属性だけ「国名」に確定する。また、このとき、この「日」、「米」の品詞情報を名詞から固有名詞に書き換える。他方、テキストB3中の「米」の属性「国名？」は消去する。こうして属性決定手段15は、図6の入力テキストに対して、図10のような解析結果を生成する。

【0050】この解析結果に基づき、品詞・属性付き単語インデックス作成手段17は、図11に示すような情報を品詞・属性付き単語インデックス18に登録する。

20 【0051】ユーザは、米国を意味する「米」を含むテキストを検索したいとする。このときユーザは、単語「米」と共にその属性である「国名」を指定した属性付き検索キー「米\$国名」を入力する。ここで、\$は単語と属性の情報とを区切るマークである。検索キー解釈手段30は、\$の前の部分を単語、\$の後ろの部分を属性の情報と解釈し、テキスト検索手段7に伝達する。テキスト検索手段7は、単語「米」をキーに図11に示す品詞・属性付き単語インデックス18を検索すると、図6のテキストB2、B3が求まる。ここで、属性を比較すると、テキストB2は合致するが、テキストB3は合致しないため、テキストB2のみを検索結果とする。

30 【0052】これに対し、図18で説明した従来のテキスト検索装置では、属性による絞り込みができないため、テキストB2だけでなくテキストB3も検索結果に含まれることになり、検索過剰となる。

【0053】上記の例において、ユーザが、属性付き検索キー「米\$国名」の代わりに、単語「米」と共にその品詞である「固有名詞」を指定した品詞付き検索キー「米@固有名詞」を入力した場合には、図1の実施例と同様の動作が行われる。また、ユーザが、単語「米」と共にその品詞である「固有名詞」とその属性である「国名」を指定した品詞・属性付き検索キー「米@固有名詞\$国名」を入力した場合には、品詞および属性の双方が合致する単語を含むテキストが検索される。

50 【0054】図12は本発明の第3の実施例のブロック図である。この例のテキスト検索システムは、テキスト記憶手段1、テキスト入力手段2、形態素解析手段3、属性追加手段100、単語展開手段19、展開知識記憶手段20および品詞・属性付き単語インデックス作成手

段 17 から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス 18、検索キー入力手段 6、検索キー解釈手段 30、テキスト検索手段 7 および検索結果記憶手段 8 から構成されるテキスト検索装置とを有している。

【0055】本実施例は、各単語の品詞の情報に加え、品詞以外の属性を単語に付加すると共に、単語を同じ意味を示す異なる表記の単語に正しく展開することにより、検索余剰を抑制しつつ、検索漏れを少なくするようにしたものである。

【0056】以下、インデックス作成装置とテキスト検索装置とに分けて説明する。

【0057】○インデックス作成装置

テキスト記憶手段 1 には、検索対象となる複数のテキストが記憶されている。テキスト入力手段 2 は、テキスト記憶手段 1 から一つずつテキストを読み込み、形態素解析手段 3 に入力する。形態素解析手段 3 は、入力されたテキストについて形態素解析を実行する。ここでの形態素解析は、入力されたテキストを単語辞書を用いて単語単位に認定し、かつ各単語の品詞を認定する解析処理である。続いて属性追加手段 100 が、図 5 の実施例と同様に、単語に対し品詞以外の属性を決定して付与する。

【0058】次に単語展開手段 19 は、属性追加手段 100 までの処理で得られたテキスト中の単語及び品詞並びに属性に基づき、展開知識記憶手段 20 に記憶された知識を参照して、単語を同じ意味の異なる表記の単語に展開する。展開知識記憶手段 20 には、単語の展開に関する知識として、その品詞、属性との関係でその同義語や概念語の辞書情報などが格納されている。単語の展開は、解析された単語の品詞や属性に基づき行われるため、例えば単語「米」はその品詞が固有名詞、またはその属性が「国名」である場合に限り、「アメリカ」等に展開される。

【0059】品詞・属性付き単語インデックス作成手段 17 は、単語展開手段 19 までの結果に基づき、形態素解析手段 3 で検出された単語及び品詞並びに属性追加手段 100 で付与された属性とその単語が出現するテキストの情報と、単語展開手段 19 で展開して得られた単語及び展開元の単語の品詞並びに付与された属性とその単語が出現するテキストの情報を、品詞・属性付き単語インデックス 18 に登録する。展開して得られた単語についても登録している点が、図 5 の実施例との相違点である。

【0060】○テキスト検索装置

検索キー入力手段 6 は、ユーザが指定した検索キーを入力する。検索キーは、従来は単語のみで構成されていたが、本実施例では、単語に加えてその品詞や属性を指定することができる。検索キー解釈手段 30 は、検索キーが入力されると、それを解釈して、単語と品詞と属性の

情報とを認識し、テキスト検索手段 7 に伝達する。

【0061】テキスト検索手段 7 は、検索キー解釈手段 30 から単語と品詞の情報が渡されると、単語と品詞の情報の双方を満たす単語を、品詞・属性付き単語インデックス 18 から検索し、単語と属性の情報が渡されると、単語と属性の情報の双方を満たす単語を、品詞・属性付き単語インデックス 18 から検索し、単語と品詞と属性の情報が渡されると、単語と品詞と属性の情報の全てを満たす単語を、品詞・属性付き単語インデックス 18 から検索する。そして、それに対応するテキストの情報を検索結果記憶手段 8 に格納する。検索結果記憶手段 8 に検索結果を格納する代わりに、またはそれと共に、図示しない表示装置やプリンタに検索結果を出力するようにしても良い。

【0062】品詞や属性の付いた検索キーでなく、従来と同じく、単語のみからなる検索キーによる検索も可能である。この場合、テキスト検索手段 7 は品詞・属性付き単語インデックス 17 の品詞、属性の情報を無視し、従来と同様の検索を行う。

【0063】また、展開された単語を除外した検索も可能である。この場合、検索キーで、展開情報を検索しない旨を指定する。指定の形式としては、検索キーの最後に所定のマーク（例えば%）を付すことが考えられる。展開された単語を除外した検索の指定は、単語のみからなる検索キー、品詞付き検索キー、属性付き検索キー、品詞・属性付き検索キーの何れでも可能である。

【0064】次に、具体例を挙げて本実施例の動作を説明する。

【0065】テキスト記憶手段 1 に記憶されているテキストが、図 6 に示すテキスト B1、B2、B3 であるとする。このテキスト B1、B2、B3 について、形態素解析手段 3 および属性追加手段 100 において、図 5 の実施例と同様の処理が施されることにより、図 6 の入力テキストに対して、図 10 のような解析結果が生成される。

【0066】展開知識記憶手段 20 は、「日」（固有名詞：国名）が「日本」（固有名詞：国名）と同義であること、「米」（固有名詞：国名）が「アメリカ」（固有名詞：国名）および「米国」（固有名詞：国名）と同義であることを、辞書情報として格納している。単語展開手段 19 は、この辞書情報を用いて、テキスト B2 内の「日」（固有名詞：国名）を「日本」（固有名詞：国名）に、「米」（固有名詞：国名）を「アメリカ」（固有名詞：国名）および「米国」（固有名詞：国名）に、それぞれ展開する。

【0067】品詞・属性付き単語インデックス作成手段 17 は、図 13 に示すように、図 10 の単語列に加えて、「日本」（固有名詞：国名）、「アメリカ」（固有名詞：国名）および「米国」（固有名詞：国名）も品詞・属性付き単語インデックス 18 に登録する。このと

き、展開して追加した語については、実際にはテキストに出現していないので、例えば属性欄に、その旨（展開）を記録しておく。

【0068】ユーザが、「アメリカ」を含むテキストを検索したいとする。また、単語のみ指定し、その品詞、属性は指定しなかったとする。検索キー解釈手段30は、検索キーの単語「アメリカ」をテキスト検索手段7に渡し、テキスト検索手段7は、「アメリカ」をキーに図13の品詞・属性付き単語インデックス18を検索する。その結果、テキストB1およびテキストB2を得ることができる。つまり、「アメリカ」を意味する「米」を含むテキストB2も得ることができる。他方、「コメ」を意味する「米」を含むテキストB3は検索されない。こうして、検索余剰を抑制しつつ、検索漏れを少なくすることができる。

【0069】これに対し、従来の単語の見出しだけをキーとした検索装置では、テキストB1しかヒットしない。つまり、「アメリカ」を意味する「米」も検索対象としたい場合に検索漏れとなる。また、検索漏れを防ぐために、ユーザ自身が「アメリカ」だけでなく、それを「米」に同義語展開して従来の検索装置で検索すると、テキストB1～B3の全てがヒットしてしまい、検索過剰となる。

【0070】なお、「米」を含むテキストは検索対象とせず、「アメリカ」だけを検索したい場合は、「アメリカ%」のように指定する。この場合、テキスト検索手段7は、検索結果からテキストB2を除外し、テキストB1だけを出力する。

【0071】その他、図5の実施例と同様に、単語の品詞、属性を指定した検索も可能である。

【0072】別の具体例を挙げて本実施例の動作を説明する。

【0073】テキスト記憶手段1に格納されているテキストC1「…自社各党は…」が、テキスト入力手段2で入力されたとする。

【0074】形態素解析手段3は、テキストC1を形態素解析する。名詞「社」と「さ」の文法非接続性から形態素解析に失敗し、「…自（名詞）／社（名詞）／さ（未知語）／各党（名詞）／は（助詞）／…」のように、「さ」の部分が未知語という結果が得られたとする。

【0075】属性記憶手段14に、「自」、「社」、「さ」という単語は政党名であるという辞書情報が格納されていると、属性付与手段13は、テキストC1に対して、「…自（名詞：政党名？）／社（名詞：政党名？）／さ（未知語：政党名？）／各党（名詞）／は（助詞）／…」のように属性を付与する。

【0076】ルール記憶手段16に、「属性情報「政党名？」が連続（近接）するパターンであれば、「政党名？」は「政党名」である」というルールが格納されてい

ると、属性決定手段15は、テキストC1に対して、「…自（名詞：政党名）／社（名詞：政党名）／さ（未知語：政党名）／各党（名詞）／は（助詞）／…」のように属性を決定する。

【0077】展開知識記憶手段20は、「自」（固有名詞：政党名）が「自民党」（固有名詞：政党名）および「自民」（固有名詞：政党名）と同義であること、「社」（固有名詞：政党名）が「社会党」（固有名詞：政党名）および「社会」（固有名詞：政党名）と同義であること、「さ」（固有名詞：政党名）が「さきがけ」（固有名詞：政党名）と同義であることを、辞書情報として格納している。

【0078】単語展開手段19は、この辞書情報を用いて、テキストC1内の「自」（固有名詞：政党名）を「自民党」（固有名詞：政党名）および「自民」（固有名詞：政党名）に、「社」（固有名詞：政党名）を「社会党」（固有名詞：政党名）および「社会」（固有名詞：政党名）に、「さ」（固有名詞：政党名）を「さきがけ」（固有名詞：政党名）に、それぞれ展開する。

【0079】品詞・属性付き単語インデックス作成手段17は、テキストC1内の単語に加えて、「自民党」（固有名詞：政党名）、「自民」（固有名詞：政党名）、「社会党」（固有名詞：政党名）、「さきがけ」（固有名詞：政党名）を、品詞・属性付き単語インデックス18に登録する。

【0080】この結果、ユーザが、「自民党」、「自民\$政党名」、「社会党」、「社会\$政党名」、「さきがけ」といった検索キーを入力した場合、これらの単語はテキスト中に実在しないものの、テキストC1を検索することができる。

【0081】図14は本発明の第4の実施例のブロック図である。この例のテキスト検索システムは、テキスト記憶手段1、テキスト入力手段2、形態素解析手段3、属性追加手段100および品詞・属性付き単語インデックス作成手段17から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス18、検索キー入力手段6、検索キー解釈手段30、テキスト検索手段7、検索結果記憶手段8、品詞・属性候補検索手段33、品詞・属性候補記憶手段34、品詞・属性候補表示手段35および品詞・属性指定手段36から構成されるテキスト検索装置とを有している。

【0082】本実施例は、図5で説明した実施例とテキスト検索装置の構成が相違している。インデックス作成装置の構成と動作は図5の実施例と同じなので、以下、テキスト検索装置についてのみ説明する。

【0083】○テキスト検索装置

検索キー入力手段6は、ユーザが指定した検索キーを入力する。検索キーには、図5の実施例と同様に、単語に加えてその品詞や属性を指定することができる。品詞や

属性が指定された場合の動作は図5の実施例と同じである。単語のみの検索キーが入力された場合の動作は、以下のように図5の実施例と相違する。

【0084】検索キー入力手段6から入力された検索キーは、検索キー解釈手段30および品詞・属性候補検索手段33に送られる。

【0085】品詞・属性候補記憶手段34は、多義性のある単語などについて、単語の見出しとその品詞・属性候補を一つまたは複数記憶する辞書である。品詞・属性候補検索手段33は、入力された検索キーが単語のみから成る場合、その単語の品詞・属性候補を品詞・属性候補記憶手段34から検索する。検索された品詞・属性候補の一覧は、品詞・属性候補表示手段35によって図示しないディスプレイなどに表示され、その選択がユーザに促される。品詞・属性指定手段36は品詞・属性候補表示手段35で表示された品詞・属性候補の一覧の中からユーザが意図する品詞・属性を指定するための手段であり、例えばキーボードによる指定やマウスによる指定を受け付ける。品詞・属性候補の一覧の中から指定（選択）された品詞・属性は、検索キー解釈手段30に送ら

れる。

【0086】検索キー解釈手段30は、検索キー入力手段6から入力された検索キーが単語のみからなる場合、品詞・属性指定手段36から送られてくる品詞または属性を、検索キーと共にテキスト検索手段7に送る。テキスト検索手段7は、検索キー解釈手段30から単語と品詞または属性の情報が渡されると、単語と品詞または属性の情報の双方を満たす単語を、品詞・属性付き単語インデックス18から検索し、それに対応するテキストの情報を検索結果記憶手段8に格納する。検索結果記憶手段8に検索結果を格納する代わりに、またそれに加えて、図示しない表示装置やプリンタに検索結果を出力するようにしても良い。

【0087】次に、具体例を挙げて本実施例のテキスト検索装置の動作を説明する。

【0088】検索キー入力手段6で、例えば単語「米」だけを含む検索キーが入力されると、品詞・属性候補検索手段33は「米」で品詞・属性候補記憶手段34を検索する。品詞・属性候補記憶手段34には、「米」の品詞・属性として一般名詞と固有名詞（国名）があることが記述されている。品詞・属性候補検索手段33は、「米」についてのこれら二つの品詞・属性を検索すると、品詞・属性候補表示手段35によって、これらを図15のようにメニュー形式でユーザに提示する。ユーザは、検索したい語が国名を意味する「米」であれば、品詞・属性指定手段36によって「米（国名）」を選択すれば、検索キー解釈手段30に、属性情報として（国名）が渡される。また、ユーザは、検索したい語が「コメ」を意味する「米」であれば、「米（一般名詞）」を選択すれば、検索キー解釈手段30に、品詞情報として

（一般名詞）が渡される。テキスト検索手段7は、これらの品詞・属性を考慮してテキスト検索を実行する。

【0089】このように本実施例では、ユーザは検索キーを入力する段階で品詞・属性を必ずしも入力する必要がない。多義性のある語などが検索キーとなった場合に、本実施例のテキスト検索装置は、品詞・属性をメニュー形式でユーザに提示し選択させるからである。これにより、ユーザの検索意図に合った検索を実行できる。また、品詞・属性の候補の一覧が表示されるため、ユーザが自ら品詞・属性を考えて指定する方法に比べて、より初心者に向いている。

【0090】以上、本発明を幾つかの実施例を挙げて説明したが、本発明は以上の実施例にのみ限定されず、その他各種の付加変更が可能である。例えば、図16に示すように、図12の実施例で用いたインデックス作成装置と、図14の実施例で用いたテキスト検索装置とを組み合わせた実施例（第5の実施例）も考えられる。また、図1の実施例で用いたインデックス作成装置と、図14の実施例で用いたテキスト検索装置を品詞に限定した装置とを組み合わせた実施例（第6の実施例）も考えられる。

【0091】図17は本発明を適用したコンピュータシステムの一例を示すブロック図である。この例のコンピュータシステムは、CPUや主記憶等を有するコンピュータ101と、キーボード102と、表示装置103と、磁気ディスク装置等の記憶装置104と、記録媒体105とから構成される。記憶装置104は、前述した各実施例におけるテキスト記憶手段1や検索結果記憶手段8等として使用される。記録媒体105は、磁気ディスク、CD-ROM、半導体メモリ等の機械読み取り可能な記録媒体であり、ここに記録されたプログラムはコンピュータ101に読み取られ、コンピュータ101の動作を制御することにより、コンピュータ101上に、前述した各実施例における構成要素を実現する。

【0092】即ち、図1の実施例にあっては、コンピュータ101上に、テキスト入力手段2、形態素解析手段3および品詞付き単語インデックス作成手段11から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞付き単語インデックス12、検索キー入力手段6、検索キー解釈手段30およびテキスト検索手段7から構成されるテキスト検索装置とを実現する。

【0093】また、図5の実施例にあっては、コンピュータ101上に、テキスト入力手段2、形態素解析手段3、属性追加手段100および品詞・属性付き単語インデックス作成手段17から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス18、検索キー入力手段6、検索キー解釈手段30およびテキスト検索手段7から構成されるテキスト検索装置とを実現する。

【0094】また、図12の実施例にあっては、コンピュータ101上に、テキスト入力手段2、形態素解析手段3、属性追加手段100、単語展開手段19、展開知識記憶手段20および品詞・属性付き単語インデックス作成手段17から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス18、検索キー入力手段6、検索キー解釈手段30およびテキスト検索手段7から構成されるテキスト検索装置とを実現する。

【0095】また、図14の実施例にあっては、コンピュータ101上に、テキスト入力手段2、形態素解析手段3、属性追加手段100および品詞・属性付き単語インデックス作成手段17から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス18、検索キー入力手段6、検索キー解釈手段30、テキスト検索手段7、品詞・属性候補検索手段33、品詞・属性候補記憶手段34、品詞・属性候補表示手段35および品詞・属性指定手段36から構成されるテキスト検索装置とを実現する。

【0096】また、図16の実施例にあっては、コンピュータ101上に、テキスト入力手段2、形態素解析手段3、属性追加手段100、単語展開手段19、展開知識記憶手段20および品詞・属性付き単語インデックス作成手段17から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞・属性付き単語インデックス18、検索キー入力手段6、検索キー解釈手段30、テキスト検索手段7、品詞・属性候補検索手段33、品詞・属性候補記憶手段34、品詞・属性候補表示手段35および品詞・属性指定手段36から構成されるテキスト検索装置とを実現する。

【0097】さらに、図1の実施例で用いたインデックス作成装置と図14の実施例で用いたテキスト検索装置を品詞に限定した装置とを組み合わせた実施例（第6の実施例）にあっては、コンピュータ101上に、テキスト入力手段2、形態素解析手段3および品詞付き単語インデックス作成手段11から構成されるインデックス作成装置と、このインデックス作成装置によって作成された品詞付き単語インデックス12、検索キー入力手段6、検索キー解釈手段30、テキスト検索手段7、品詞候補検索手段（33）、品詞候補記憶手段（34）、品詞候補表示手段（35）および品詞指定手段（36）から構成されるテキスト検索装置とを実現する。

【0098】

【発明の効果】以上説明したように本発明によれば以下のような効果を得ることができる。

【0099】単語インデックスを作成する際に、従来のように単語の見出しとその単語を含むテキストの情報だけでなく、その単語の品詞、属性の情報を登録することによって、品詞、属性を検索条件とした検索が可能とな

り、特に多義性のある単語を検索キーとしたときの検索過剰を抑制することができる。

【0100】また、形態素解析で得られた単語の品詞や属性追加手段で追加された属性に基づいて多義性のある語を展開して登録することにより、多義性のある語を正しく展開した単語インデックスが生成できる。これにより、検索キーと異表記でも同義の語を検索することができて検索漏れを低減でき、かつ、ユーザ自身が検索キーを同義語に展開して検索する場合に比べて検索過剰を抑えることができる。

【図面の簡単な説明】

【図1】本発明の第1の実施例のブロック図である。

【図2】入力テキストの一例を示す図である。

【図3】図2の入力テキストに対する形態素解析例を示す図である。

【図4】品詞付き単語インデックスの例を示す図である。

【図5】本発明の第2の実施例のブロック図である。

【図6】入力テキストの別の例を示す図である。

【図7】図6の入力テキストに対する形態素解析結果の一例を示す図である。

【図8】属性記憶手段の記憶内容の例を示す図である。

【図9】図7の形態素解析結果中の単語に対して属性付与手段が属性を付与した例を示す図である。

【図10】図7の形態素解析結果中の単語に対して最終的に付与された属性例を示す図である。

【図11】品詞・属性付き単語インデックスの例を示す図である。

【図12】本発明の第3の実施例のブロック図である。

【図13】展開語をも含む品詞・属性付き単語インデックスの例を示す図である。

【図14】本発明の第4の実施例のブロック図である。

【図15】品詞・属性候補の一覧をユーザに提示するメニューの例を示す図である。

【図16】本発明の第5の実施例のブロック図である。

【図17】本発明を適用したコンピュータシステムの一例を示すブロック図である。

【図18】従来のテキスト検索システムの構成を示すブロック図である。

【図19】従来の単語インデックスの内容例を示す図である。

【符号の説明】

- 1…テキスト記憶手段
- 2…テキスト入力手段
- 3…形態素解析手段
- 4…単語インデックス作成手段
- 5…単語インデックス
- 6…検索キー入力手段
- 7…テキスト検索手段
- 8…検索結果記憶手段

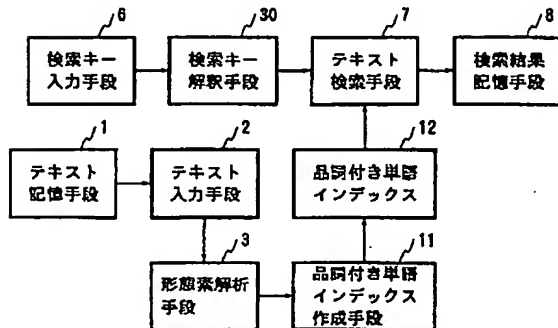
- 11…品詞付き単語インデックス作成手段
 12…品詞付き単語インデックス
 13…属性付与手段
 14…属性記憶手段
 15…属性決定手段
 16…ルール記憶手段
 17…品詞・属性付き単語インデックス作成手段
 18…品詞・属性付き単語インデックス

- * 19…単語展開手段
 20…展開知識記憶手段
 30…検索キー解釈手段
 33…品詞・属性候補検索手段
 34…品詞・属性候補記憶手段
 35…品詞・属性候補表示手段
 36…品詞・属性指定手段

*

【図1】

【図2】



テキストA1:「…大分県は…」
 テキストA2:「…大分昔に」

【図6】

テキストB1:「…アメリカは、…」
 テキストB2:「…日米間の貿易…」
 テキストB3:「…新潟産の米を…」

【図8】

【図3】

テキストA1:「…大分(固有名詞)県(固有名詞接辞)
 は(助詞)…」

テキストA2:「…大分(副詞)昔(名詞)に(助詞)」

アメリカ: 固有名詞: 国名
 米: 固有名詞: 国名
 日: 固有名詞: 国名
 新潟: 固有名詞: 県名
 大分: 固有名詞: 県名
 …

【図7】

【図15】

テキストB1:「…/アメリカ(固有名詞)/は(助詞)、…」

テキストB2:「…/日(名詞)/米(名詞)/間(名詞)/
 の(助詞)/貿易(サ変名詞)/…」

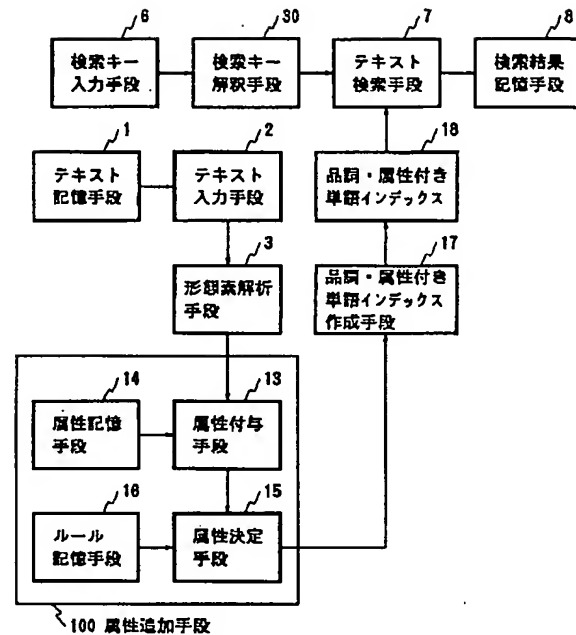
テキストB3:「…/新潟(固有名詞)/産(名詞)/の(助詞)/
 米(名詞)/を(助詞)…」

米(国名)
 米(一般名詞)

【図4】

見出し	品詞	テキスト識別子
は	助詞	A1
に	助詞	A2
昔	名詞	A2
大分	固有名詞	A1
	副詞	A2
県	固有名詞	A1
...	接辞	

【図5】



【図9】

テキストB1:「…/アメリカ(固有名詞:国名)/は(助詞)、…」
 テキストB2:「…/日(名詞:国名?)/米(名詞:国名?)/
 間(名詞)/の(助詞)/貿易(サ変名詞)/…」
 テキストB3:「…/新潟(固有名詞:県名)/産(名詞)/
 の(助詞)/米(名詞:国名?)/を(助詞)…」

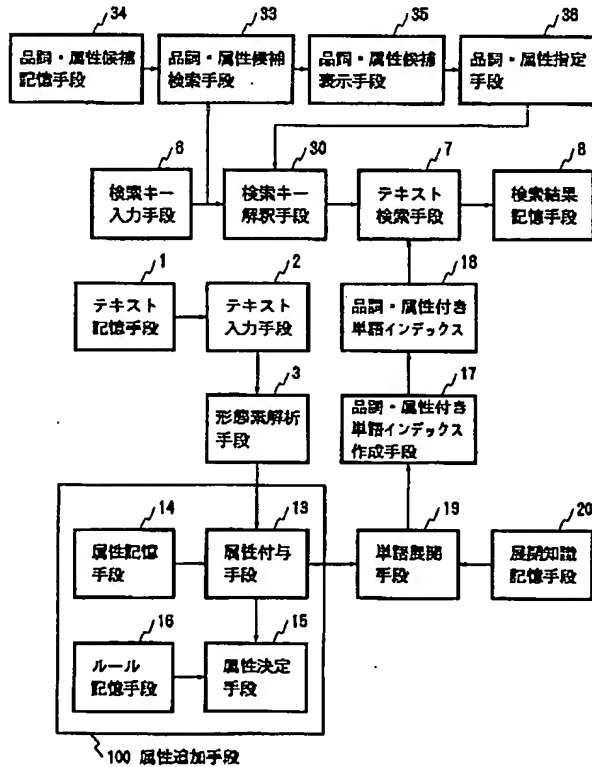
【図10】

テキストB1:「…/アメリカ(固有名詞:国名)/は(助詞)、…」
 テキストB2:「…/日(固有名詞:国名)/米(固有名詞:国名)/
 間(名詞)/の(助詞)/貿易(サ変名詞)/…」
 テキストB3:「…/新潟(固有名詞:県名)/産(名詞)/
 の(助詞)/米(名詞)/を(助詞)…」

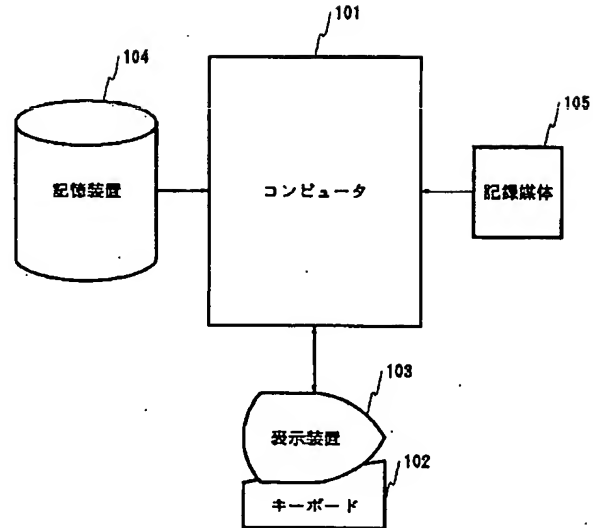
【図13】

見出し	品詞	属性	テキスト識別子
の	助詞		B2、B3
は	助詞		B1
を	助詞		B3
アメリカ	固有名詞	国名	B1
	固有名詞	国名、展開	B2
間	名詞		B2
産	名詞		B3
新潟	固有名詞	県名	B3
日	固有名詞	国名	B2
米	固有名詞	国名	B2
	名詞		B3
貿易	サ変名詞		B2
日本	固有名詞	国名、展開	B2
米国	固有名詞	国名、展開	B1、B2
...			

【図16】



【図17】



【図18】

